Psychological Medicine

Determinants of patient-reported outcome trajectories and symptomatic recovery in Improving Access to Psychological Treatment (IAPT) services

--Manuscript Draft--

Manuscript Number:	PSM-D-20-00715R4
Full Title:	Determinants of patient-reported outcome trajectories and symptomatic recovery in Improving Access to Psychological Treatment (IAPT) services
Article Type:	Original Article
Corresponding Author:	Jan Stochl, Ph.D. University of Cambridge Cambridge, UNITED KINGDOM
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	University of Cambridge
Corresponding Author's Secondary Institution:	
First Author:	Jan Stochl, PhD
First Author Secondary Information:	
Order of Authors:	Jan Stochl, PhD
	Emma Soneson, MPhil
	Freya Stuart
	Jessica Fritz, MSc
	Annabel E.L. Walsh
	Tim Croudace, PhD
	Joanne Hodgekins, PhD
	Ushma Patel
	Debra A. Russo
	Clare Knight
	Peter B. Jones, PhD
	Jesus Perez, PhD
Order of Authors Secondary Information:	
Manuscript Region of Origin:	UNITED KINGDOM
Abstract:	Background: Despite evidence for the general effectiveness of psychological therapies, there exists substantial heterogeneity in patient outcomes. We aimed to identify factors associated with baseline severity of depression and anxiety symptoms, rate of symptomatic change over the course of therapy, and symptomatic recovery in a primary mental health care setting. Methods: Using data from a service evaluation involving 35,527 patients in England's psychological and wellbeing (IAPT) services, we applied latent growth models to explore which routinely-collected sociodemographic, clinical, and therapeutic variables were associated with baseline symptom severity and rate of symptomatic change. We used a multilevel logit model to determine variables associated with symptomatic recovery. Results: Being female, younger, more functionally impaired, and more socioeconomically disadvantaged was associated with higher baseline severity of both depression and anxiety symptoms. Being older, less functionally impaired, and having more severe baseline symptomatology was associated with more rapid improvement of

	both depression and anxiety symptoms (male gender and greater socioeconomic disadvantage were further associated with rate of change for depression only). Therapy intensity and appointment frequency seemed to have no correlation with rate of symptomatic improvement. Patients with lower baseline symptom severity, less functional impairment, and older age had a greater likelihood of achieving symptomatic recovery (as defined by IAPT criteria). Conclusions: We must continue to investigate how best to tailor psychotherapeutic interventions to fit patients' needs. Patients who begin therapy with more severe depression and/or anxiety symptoms and poorer functioning merit special attention, as these characteristics may negatively impact on recovery.

Determinants of patient-reported outcome trajectories and symptomatic recovery in Improving Access to Psychological Treatment (IAPT) services

Jan Stochl, PhD[1,2,3×*], Emma Soneson, MPhil[1×], Freya Stuart[1], Jessica Fritz, MSc[1], Annabel E. L. Walsh[4], Tim Croudace, PhD[5], Joanne Hodgekins, PhD[6], Ushma Patel[1,7], Debra A Russo[1], Clare Knight[1], Peter B Jones, PhD[1,2,7+], Jesus Perez, PhD[1,2,7+]

Financial support:
This paper presents independent research funded by the National Institute for Health Research (NIHR) under its Programme Grants for Applied Research Programme (Reference Number RP-PG-0616-20003). JS, JP, and PBJ received support from the NIHR Applied Research Collaboration (ARC) East of England (EoE) at the Cambridgeshire and Peterborough NHS Foundation Trust. ES is supported by a Gates Cambridge Scholarship. JF is funded by the Medical Research Council Doctoral Training/Sackler Fund and the Pinsent Darwin Fund. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, the Department of Health and Social Care, the MRC or the Gates Cambridge Trust.

Word count: 4,665

× Joint first authors
+ Joint last authors

[1] Department of Psychiatry, University of Cambridge, Cambridge, UK
[2] National Institute for Health Research Applied Research Collaboration (ARC) East of England (EoE), Cambridge, UK
[3] Department of Kinanthropology, Charles University, Prague, Czechia
[4] Institution of Psychiatry, Psychology and Neuroscience, King's College London, UK
[5] School of Health Sciences, University of Dundee, Dundee, UK
[6] Norwich Medical School, University of East Anglia, Norwich, UK
[7] Cambridgeshire & Peterborough NHS Foundation Trust, UK

* *Address for correspondence:*
Department of Psychiatry
University of Cambridge
Cambridge Biomedical Campus
Herchel Smith Building for Brain & Mind Sciences
Cambridge, CB2 0SZ
United Kingdom
Phone: +44 07587146299
Email: js883@cam.ac.uk

Introduction

Evidence abounds for the effectiveness of psychological therapies in treating a wide range of mental health problems (Barth et al., 2013; Cristea et al., 2017; Cuijpers, Cristea, Karyotaki, Reijnders, & Huibers, 2016; Pim Cuijpers et al., 2014; Lambert, 2013). However, it is widely acknowledged that treatment outcomes vary greatly between individuals with a significant proportion not responding at all (Van, Dekker, Peen, Van Aalst, & Schoevers, 2008; Van, Schoevers, et al., 2008). The ability to explain why individuals respond differently to therapy provides important supplementary information to advance our understanding of 'what works for whom' (Stuart A. Green et al., 2015; Van, Dekker, et al., 2008). Characterising the basis of heterogeneity in baseline symptom severity, rates of symptomatic change during therapy, and treatment outcomes enables us to identify variables related to treatment success and thus may constitute a step toward more personalised care. Furthermore, provision of appropriate, tailored treatment enables efficient allocation of limited mental health resources.

Much effort has gone into understanding the reasons for variation in therapy response. At the patient level alone, more than 200 factors have been proposed to potentially influence therapy outcomes (Norcross & Wampold, 2011). These variables include sociodemographic characteristics (e.g. age (Amati, Banks, Greenfield, & Green, 2018; Marttunen, Välikoski, Lindfors, Laaksonen, & Knekt, 2008; Robinson, Kellett, & Delgadillo, 2020; Wolitzky-Taylor, Arch, Rosenfield, & Craske, 2012), gender (Amati et al., 2018; P. Cuijpers et al., 2014; Wolitzky-Taylor et al., 2012), socioeconomic status (Amati et al., 2018; S. A. Green et al., 2015; Marttunen et al., 2008), and ethnicity (Amati et al., 2018; Stuart A. Green et al., 2015; Robinson et al., 2020; Saxon, Firth, &

Barkham, 2017; Wolitzky-Taylor et al., 2012)), mental health-related clinical variables (e.g. pre-treatment disorder severity (Amati et al., 2018; Stuart A. Green et al., 2015; A. Gyani, R. Shafran, R. Layard, & D. M. Clark, 2013; Marttunen et al., 2008; Robinson et al., 2020; Saxon et al., 2017; Van, Dekker, et al., 2008; Wolitzky-Taylor et al., 2012), comorbidities (Amati et al., 2018; Goddard, Wingrove, & Moran, 2015; Marttunen et al., 2008; Wolitzky-Taylor et al., 2012), and psychiatric history (Marttunen et al., 2008)), social functioning and support (Amati et al., 2018; Lindfors, Ojanen, Jääskeläinen, & Knekt, 2014; Wang, Mann, Lloyd-Evans, Ma, & Johnson, 2018), and personality traits (Bucher, Suzuki, & Samuel, 2019; Maarit A. Laaksonen, Knekt, & Lindfors, 2013; M. A. Laaksonen, Knekt, Sares-Jäske, & Lindfors, 2013). Outcomes further vary by treatment variables (e.g. therapy modality (Amati et al., 2018; Alex Gyani, Roz Shafran, Richard Layard, & David M. Clark, 2013; Marttunen et al., 2008), number of sessions attended (Amati et al., 2018; Alex Gyani et al., 2013; Norcross & Wampold, 2011) and missed (Amati et al., 2018; Stuart A. Green et al., 2015; Van, Dekker, et al., 2008), time waited to start treatment (Clark et al., 2018), therapy setting (Amati et al., 2018), frequency of sessions (Tiemens et al., 2019), therapeutic alliance (Del Re, Flückiger, Horvath, Symonds, & Wampold, 2012; Horvath, Del Re, Flückiger, & Symonds, 2011), treatment engagement (Dixon, Holoshitz, & Nossel, 2016), and patient expectations of therapy outcome (Porter & Chambless, 2015)) as well as on therapist characteristics and experience (Amati et al., 2018; Alex Gyani et al., 2013; Nissen-Lie, Monsen, Ulleberg, & Rønnestad, 2013). Whilst some of these factors influence treatment outcome in a consistent direction (good therapeutic alliance, for example, consistently leads to more

positive outcomes (Del Re et al., 2012; Horvath et al., 2011)), several show inconclusive evidence regarding direction of effect (Amati et al., 2018).

Individuals may further vary in their progression through therapy, leading to heterogeneity in treatment response trajectories (Stuart A. Green et al., 2015). These trajectories are informative for understanding progress as well as predicting outcomes (Comninos & Grenyer, 2007; W. Lutz et al., 2014; W. Lutz, Stulz, & Kock, 2009; Stulz, Lutz, Leach, Lucock, & Barkham, 2007). Research has aimed to characterise differences in treatment response trajectories, largely through class-based approaches that classify sub-groups of patients with homogeneous treatment response trajectories and determine predictors of group membership. These studies have identified varying number of such sub-groups for a wide range of disorders, including depression (Cuijpers, Van Lier, Van Straten, & Donker, 2005; Gunn et al., 2013; W. Lutz et al., 2009; Sunderland, Wong, Hilvert-Bruce, & Andrews, 2012), anxiety (Sunderland et al., 2012), panic disorder (W. Lutz et al., 2014), post-traumatic stress disorder (PTSD) (Elliott, Biddle, Hawthorne, Forbes, & Creamer, 2005; Stein, Dickstein, Schuster, Litz, & Resick, 2012), and first-episode psychosis (Hodgekins et al., 2015) .

Previous studies of both therapy outcomes and treatment response trajectories in psychological therapy have some limitations. First, with few exceptions (Ali et al., 2014; Flückiger, Grosse Holtforth, Znoj, Caspar, & Wampold, 2013; Green, Barkham, Kellett, & Saxon, 2014; Wolfgang Lutz, Martinovich, Lyons, Leon, & Stiles, 2007; Saxon et al., 2017), they have not accounted for the structure of longitudinal data in which patients are nested under individual therapists. Failure to take into account this hierarchical data structure can result in biased statistical inferences (Stochl et al., 2014). Second, many

of the previous studies have relied upon specialised, non-routine variables (e.g. measures of therapeutic alliance). Whilst these provide valuable insights, there is a practical need for easily identifiable, routinely-collected variables, including patients' sociodemographic characteristics and clinical features (Van, Dekker, et al., 2008).

These limitations highlight a need for robust studies that use appropriately complex multilevel models to cope with hierarchical and longitudinal dependencies, as well as convenient (i.e. readily available in IAPT's routinely-collected variables) and practical (i.e. relevant to outcomes of interest) variables to explain heterogeneity in treatment response trajectories and symptomatic recovery. In this exploratory analysis, we aimed to identify variables that are associated with (1) baseline symptom severity, (2) rate of symptomatic change, and (3) symptomatic recovery of patients receiving psychological therapy in England's Improving Access to Psychological Therapies (IAPT) primary mental health care setting.

Methods

Setting

The IAPT programme in England began in 2008 with a direct objective to increase public access to National Institute for Health and Care Excellence (NICE) approved psychological therapies for depression and anxiety. IAPT currently assesses over 1,300,000 people annually and delivers therapy to approximately 550,000. The programme offers low- (step 2) and high-intensity (step 3) treatment. Low-intensity approaches include guided self-help, psychoeducation, computerised CBT, behavioural activation, and structured group activity programmes (Clark, 2018). In high-intensity

services, face-to-face cognitive–behavioural therapy (CBT) is the predominant approach, although there is a wider range of recommended treatments (e.g. eye movement desensitisation and reprocessing (EMDR), interpersonal psychotherapy (IPT), counselling for depression, compassion-focused therapy (CFT), and integrative counselling). On average, patients receive 7 sessions over a period of 3-4 months.

At each session, therapists assess depression and anxiety symptoms using the 9-item Patient Health Questionnaire (PHQ-9) (Kroenke, Spitzer, & Williams, 2001) and the 7-item Generalized Anxiety Disorder assessment (GAD-7) (Spitzer, Kroenke, Williams, & Lowe, 2006), respectively. IAPT services adopted these scales nationally because of their good psychometric properties (Cameron, Crawford, Lawton, & Reid, 2008; Titov et al., 2011) and brevity, and they use them to monitor improvement and recovery rates. Total scores are computed as sum scores of items (response categories: 0=Not at all; 1=Several days, 2=More than half the days; 3=Nearly every day). PHQ-9 scores range from 0 (no depression) to 27 (severe depression), whilst GAD-7 scores range between 0 (no anxiety) and 21 (severe anxiety). In IAPT, individuals are described at ‘caseness’, if they score above the clinical cut-off for depression (PHQ-9 ≥ 10) (Manea, Gilbody, & McMillan, 2012) and/or anxiety (GAD-7 ≥ 8).

Participants

The primary sample consisted of 35,527 individuals across Cambridgeshire and Peterborough NHS Foundation Trust and Sussex Partnership NHS Foundation Trust who accessed IAPT services between February 2018 and December 2018. We

excluded 6,717 individuals deemed not suitable for the service after initial assessment and those with no longitudinal data (i.e. who only attended one appointment). The sample analysed for determinants of baseline symptom severity and rate of symptomatic change consisted of 28,810 individuals (Table 1).

To assess which variables were associated with symptomatic recovery, we analysed therapy outcomes for a subsample 8,114 individuals comprising those patients who: a) had non-missing values for both PHQ-9 and GAD-7 scores at first and last assessment (n=27,835); b) were considered at 'caseness' at initial assessment (i.e. PHQ-9 and/or GAD-7 score was above clinical cut-off, n=20,959); c) had completed treatment (in any number of sessions, n=10,308[1]); and d) had non-missing values on all variables hypothesised to be associated with outcomes (n=8,114). Over 67% (5,438) individuals achieved symptomatic recovery.

Outcomes

Our outcomes of interest were *baseline symptom severity, rate of symptomatic change,* and *symptomatic recovery*. We inferred each patient's baseline symptom severity and rate of symptomatic change from his/her growth curve (hereafter denoted as *treatment response trajectory*), which is estimated using his/her total scores on the corresponding scale across therapy appointments. *Symptomatic recovery* in IAPT is achieved when a patient who is at 'caseness' at pre-treatment has dropped below the clinical cut-off for

[1] In response to a suggestion following peer review, we have also analyzed the sample that included individuals who had not completed treatment and whose symptomatic recovery was derived from the last recorded session. The results are presented in Appendix 4.

185 *both* depression and anxiety post-treatment (i.e. PHQ-9 < 10 *and* GAD-7 < 8) (National
186 Collaborating Centre for Mental Health, 2018).

187

188 **Variables tested for association with outcomes**

189 The set of variables tested for associations with our outcomes of interest included three
190 sociodemographic variables (*gender, age,* and *socioeconomic status*), two clinical
191 variables (*baseline symptom severity* and *baseline level of functioning*), and three
192 therapeutic variables (*therapy intensity* (low- (step2) or high-intensity (step3)), *therapy*
193 *frequency* (median number of days between sessions), and the number of patients
194 treated by an individual therapist (*caseload*)). All of these variables were directly
195 obtained or derived from data routinely collected in IAPT services (see Supplementary
196 Table 1 for detailed variable list).

197 *Baseline level of functioning* was measured using the 5-item self-report Work and
198 Social Adjustment Scale (WSAS) (Mundt, Marks, Shear, & Greist, 2002), which
199 measures personal, occupational, and social functional impairment. Each item on the
200 WSAS is scored from 0 (no impairment) to 8 (severe impairment), hence total scores
201 range from 0 to 40, with higher total scores indicating more severe impairment.

202 *Socioeconomic status* was estimated using the Index of Multiple Deprivation
203 (IMD). IMD deciles are publicly available data for postcodes across the UK. Anonymised
204 data provided by IAPT, however, included only the outward area postcode. The IMD for
205 each individual was therefore estimated as a median IMD decile for the corresponding
206 outward area, with lower values representing higher deprivation.

Statistical analysis

We used growth models to estimate patients' treatment response trajectories of depression and anxiety symptoms. This modelling approach fits, for each patient, a non-linear trend for the total scores of PHQ-9 or GAD-7 over the course of therapy. We inferred values for baseline symptom severity and rate of symptomatic change from the intercept and slope of patients' treatment response trajectories, respectively. In all analyses we used Full Information Maximum Likelihood estimator to account for data missingness.

The IAPT treatment model is based on 12-20 sessions as set out in the NICE guidelines. Very few individuals attended more than 20 appointments and thus we only used the first 20 appointments for each patient. We first estimated the basic nonlinear growth model (with intercept, slope, and quadratic term as latent variables) anchored at every attended appointment (see Supplementary Figure 1).

At this stage, we investigated whether the individual treatment trajectories cluster into homogeneous classes using a growth mixture model. If few interpretable classes are found, then the variables associated with baseline symptom severity and rate of symptomatic change could be considered for these classes instead of individual trajectories. However, we did not find such classes (see Appendix 1) and thus carried out the analysis on an individual level.

Next, we reparameterised the model so that the slope represented change over the first 7 appointments rather than between the first and the second appointments as in the basic model (note that all 20 appointments are still used to estimate this model). We

chose to report the symptomatic change in the first 7 appointments because it is the average number of appointments nationally and thus represents a typical length of therapeutic intervention within IAPT (Public Health England, 2019). We detail the model in Appendix 2.

Additional analytical complexity stemmed from the multilevel structure of the data (multiple patients received therapy from the same therapist). Accounting for this analytically provides unbiased treatment response trajectories. In addition, it allowed us to assess variables associated with the average rate of symptomatic change for each therapist's patients. Additional details are provided in Appendix 3.

We then added variables considered to be related with the intercepts (baseline symptom severity) and slopes (rate of symptomatic change) of treatment response trajectories to the reparameterised model. At the patient level (the within-level part of the model), we included gender, age, socioeconomic status, and baseline level of functioning (WSAS). We included therapy intensity (low vs high) as an important covariate accounting for different therapeutic approaches applied for low and high intensity IAPT patients. Therapy frequency was considered only in the context of the slope as its association with the intercept would be conceptually non-sensical). We also explored the association between baseline symptom severity and rate of symptomatic change. At the therapist level (the between-level part of model), we examined whether caseload was associated with a) the average rate of symptomatic change and b) recovery for each therapist's patients.

We used a multilevel logit model for variables associated with the binary clinical endpoint of symptomatic recovery. The set of variables examined in relation to

symptomatic recovery was identical to those of rate of symptomatic change except that a) we included baseline symptom severity of both depression and anxiety (as recovery in IAPT requires having scores below corresponding threshold on both measures) and b) these baseline symptom severities were operationalised as total scores of PHQ-9 and GAD-7 at initial assessment (i.e. not as intercepts of treatment response trajectories). Total PHQ-9 and GAD-7 scores at baseline were moderately correlated (r=0.46), allowing inclusion of both variables in the same model. We bootstrapped the model (1000 iterations) to obtain bootstrapped confidence intervals for odds ratios. We conducted all analyses in MPlus 8.4 (Muthén & Muthén, 1998-2019) and R 3.6.3 (R Core Team, 2019).

The script for our analyses and synthetic data are available at https://osf.io/48eur/.

Results

Sample and variable descriptives

Table 1 provides the sociodemographic characteristics and basic descriptive statistics of variables hypothesised to be associated with outcomes.

---------- insert Table 1 about here ----------

Growth models

The basic nonlinear growth models fit the data well (PHQ-9: RMSEA=0.034, CFI=0.952, TLI=0.955, SRMR=0.082; GAD-7: RMSEA=0.032, CFI=0.951, TLI=0.954, SRMR=0.080). Figure 1 depicts the estimated mean trajectories for the two scales. The modelled mean baseline symptom severity (intercept) had a value of 13.5 for the PHQ-9 and 12.6 for the GAD-7. The mean slopes (-0.9 for the PHQ-9 and -0.8 for the GAD-7) reflect the average change in scores between the first and second therapy session. Such interpretation is not very informative in regard to understanding improvement over the course of the therapy. In the reparameterised growth model, the slopes represent change over the first 7 sessions. Results suggest that the average improvement across 7 therapy sessions is 4.2 points on the PHQ-9 and 4.0 points on the GAD-7 (see Appendix 2 for details).

---------- insert Figure 1 about here ----------

Variables associated with baseline symptom severity and rate of symptomatic change

Tables 2 and 3 include estimated regression coefficients (including standardised estimates) for depression and anxiety symptoms, respectively.

Variables associated with baseline symptom severity and rate of symptomatic change for depression (PHQ-9)

Patients' modelled baseline symptom severity (the within-level intercept of his/her treatment response trajectory) was significantly related to gender (females have greater baseline symptom severity), age (younger patients have greater baseline symptom severity), baseline functioning (patients with more functional impairment have greater baseline symptom severity), socioeconomic status (patients living in areas of higher deprivation have greater baseline symptom severity), and therapy intensity. The significant positive relationship between baseline symptom severity and therapy intensity confirms that patients with more severe depression symptoms tend to be assigned to high-intensity therapy. In terms of standardised coefficients, baseline symptom severity had the strongest relationship with baseline functioning scores.

Patients' rate of symptomatic change for depression (the within-level slope of patients' treatment response trajectories) was most strongly related (in terms of magnitude of impact) to their baseline depression severity. More specifically, the higher the baseline symptom severity, the faster the improvement. Additionally, the rate of symptomatic change was related to gender (males improve more rapidly), age (older patients improve more rapidly), baseline level of functioning (patients with less functional impairment improve more rapidly), and socioeconomic status (patients living in areas of higher deprivation improve more rapidly). The improvement rate was *not* significantly related to therapy frequency or intensity.

The average improvement of a particular therapist's patients (the between-level part of the model), was significantly related to that therapist's caseload, however, in an unexpected direction – a larger caseload was related to more rapid improvement.

---------- insert Table 2 about here ----------

Variables associated with baseline symptom severity and rate of symptomatic change for anxiety (GAD-7)

Results for the GAD-7 treatment response trajectories were similar to those of the PHQ-9. However, contrary to the PHQ-9 results, the relationship between baseline symptom severity and therapy intensity was at the borderline of statistical significance and was in the opposite direction than for depression. Furthermore, gender and socioeconomic status were not related to the rate of improvement for GAD-7 scores. Finally, considering the standardised coefficients, functioning has a larger effect size for anxiety than for depression (with lower impairment related to more rapid improvement), when controlling for all other variables.

---------- insert Table 3 about here ----------

Variables associated with symptomatic recovery

Table 4 shows results of a multilevel logit model with symptomatic recovery as the outcome. Baseline symptom severity, age, and baseline level of functioning were significantly related to symptomatic recovery, even when bootstrap was applied. Specifically, with each additional point on the PHQ-9 or GAD-7 at the beginning of therapy, the chances of symptomatic recovery decreases by approximately 5.8% and 6.3%, respectively. Similarly, an increase of one point on the WSAS is associated with an approximate 2.4% reduction in chances of recovery. Each additional year of age increases odds of recovery by approximately 0.8%.

Therapy intensity and frequency were significantly related to symptomatic recovery, however, bootstrapped odd ratios were not and thus the results should be interpreted with caution. In addition, both variables have a relatively small effect on recovery. For example, an increase of a day in the median number of days between sessions reduced the probability of recovery by 0.7% and being assigned to high-intensity therapy lowers chances of recovery by approximately 11.2% when adjusting for all other variables (it is important to note, however, that high-intensity therapists generally see more complex and severe patients).

Socioeconomic status, gender, and therapist caseload were not significantly associated with symptomatic recovery.

The results for the sample including those individuals who had not yet completed therapy or who had dropped out but whose symptomatic recovery was derived from the last recorded session, are presented in Appendix 4. They were very similar to those presented in Table 4.

---------- insert Table 4 about here ----------

Discussion

In this study, we explored which sociodemographic, clinical, and therapeutic variables may be related to (1) baseline symptom severity, (2) rate of symptomatic change, and (3) symptomatic recovery (as defined by IAPT criteria) for patients with depression and anxiety engaging in psychological therapy in IAPT services. Importantly, our predictors were all variables routinely collected in IAPT services and included gender, age, and socioeconomic status, clinical variables (baseline depression and anxiety scores, and baseline level of functioning), and therapeutic variables (therapy intensity and median number of days between sessions (i.e. frequency)). Our multilevel approach also allowed us to test the relationship between therapists' IAPT caseload and the average rate of symptomatic change and recovery of their corresponding patients.

Treatment response trajectories

Whilst baseline severity of both depression and anxiety symptoms was significantly related to all included variables (i.e. gender, age, baseline functioning, socioeconomic status, and therapy intensity), baseline functioning had by far the largest effect sizes (whereby greater functional impairment was associated with more severe symptomatology). Age, baseline functioning, baseline symptom severity, and therapist caseload were significantly related to the rate of change for both depression and anxiety symptoms, and gender and socioeconomic status were additionally related to rate of change for depression (but not anxiety) symptoms. Therapy characteristics such as intensity and frequency seemed to have no relation to the rate of symptomatic change.

Baseline symptom severity was the most important variable associated with rate of change for both depression and anxiety symptoms, whereby patients with more severe symptomatology at the start of therapy improved more rapidly. On average, a 1-point difference in total score on the corresponding scale between two patients at baseline results in an expected difference of 0.748 (i.e. 1 – 0.252) points on the PHQ-9 and 0.738 (i.e. 1 – 0.262) points on the GAD-7 between them at 7th session. Baseline functioning was the only other variable to show a clinically relevant effect size at the individual level (whereby greater functional impairment at the start of therapy was related to slower improvement), but only for anxiety symptoms.

Our finding that greater baseline severity of depression and anxiety symptoms was associated with more rapid symptom improvement is unusual. In general, others have found either that baseline severity negatively impacts the rate of symptomatic improvement (e.g. Sunderland and colleagues' (2012) study of online CBT for depression and anxiety disorders) or has no relationship (e.g. Comninos and Greyner's (2007) study of early rapid response in supportive-expressive dynamic psychotherapy for major depression). The importance of baseline symptom severity extends beyond symptomatic change, as Hodgekins et al. (2015) demonstrate in their finding that baseline severity of psychotic symptoms serves as a predictor of belonging to a 'poorer' trajectory of social recovery for patients with first episode psychosis.

We located only one study that reflected our finding that baseline severity was positively related to the rate of symptomatic improvement (Elliott and colleagues' (2005) study of veterans receiving treatment for PTSD). It is conceivable that help-seeking individuals with higher baseline symptom severity may be more motivated to engage

with therapy in an effort to overcome more severe symptoms ('the gift of desperation') and that increased engagement positively impacts on treatment outcomes (Dixon et al., 2016). However, this explanation is perhaps more reasonable for anxiety symptoms than depression symptoms, wherein greater severity may instead be a barrier to engagement. Furthermore, it is important to consider the potential for statistical artefacts when interpreting these effect sizes, as patients starting with more severe symptoms have greater scope for improvement.

Our finding about the negative impact of functional impairment on rate of symptomatic improvement is more consistent with the literature. For example, Lutz and colleagues' (2014) study of CBT for patients with panic disorder highlighted the importance of social functioning in predicting class membership (as characterised in part by rate of symptomatic change). Poor functioning is a well-documented barrier to symptomatic improvement. Many people with greater functional impairment are unemployed and/or have fewer social contacts, and thus are missing two key protective factors associated with positive mental health. Poor functioning is further related to barriers in engaging with therapy. For example, poorer functioning may equate to fewer resources for use in therapy or poor attendance (due to various reasons including financial or social difficulties and anxiety). Thus, interventions for those with poorer social functioning may require additional considerations, such as strategies for returning to work or building up social networks and overcoming barriers related to each of these goals (Knight et al., 2019).

At the therapist level, our finding that a larger therapist caseload was associated with more rapid improvement could indicate that more frequent application of IAPT

techniques facilitates greater therapist competency. Alternatively, this could simply indicate that more competent therapists are assigned more patients. In either case, this result should not be interpreted causally. First, the caseload variable represents only the number of *IAPT* patients seen by each therapist and is not weighted for the number of days worked in IAPT. It is possible that each therapist sees additional patients outside of IAPT, in which case the effect of *total* caseload would be unmeasured in our analyses. Second, some IAPT therapists focus on specific groups of patients (e.g. patients with long-term physical conditions), which could affect their caseload and potentially bias results; however, this specialisation applies to a relatively small group of therapists.

Symptomatic recovery

Higher chance of symptomatic recovery, as defined by IAPT criteria, was associated with lower baseline severity of depression and anxiety symptoms, lower functional impairment, and increased age, with baseline symptom severity having the greatest effect. These findings are not particularly surprising: whilst starting therapy with a higher score on the PHQ-9/GAD-7 enables more scope for improvement (hence the sensibility of its association with faster symptomatic improvement), it also implies a further distance to IAPT's recovery 'threshold'. Several other studies have found similar results in terms of baseline symptom severity (Amati et al., 2018; Marttunen et al., 2008), including two within the IAPT setting (S. A. Green et al., 2015; A. Gyani et al., 2013). It is conceivable that patients starting therapy with greater symptom severity and functional impairment have more difficulty engaging in treatment than those with less

severe symptomatology and impairment. These patients may also be at increased risk for additional psychiatric comorbidities, including psychotic experiences (Stochl et al., 2015), which could further contribute to the reduced chance of recovery (Knight et al., 2019).

In interpreting our results, it must be acknowledged that people with 'less favourable' characteristics (e.g. those with higher baseline symptom severity/functional impairment) do not necessarily benefit *less* from therapy. It is important to remember that the definition of symptomatic recovery (as routinely used in IAPT for performance monitoring purposes) is centred around *absolute improvement* (i.e. whether their symptoms were reduced beyond the recovery 'threshold') rather than *relative improvement* (i.e. the difference between baseline and final symptom severity). Hence, in order to be most informative, results about symptomatic recovery should be contextualised within our discussion of treatment response trajectory.

Strengths and limitations

The main strength of this study is the large clinical sample. Furthermore, although this is not the first study that has explored variables related to symptomatic improvement and recovery in the context of psychological therapy, we used an analytical approach that is more appropriate for the complex data structure of longitudinal outcomes nested under therapists. A further strength is the applicability to clinical settings in general, and to the IAPT setting in particular, as we used routinely-collected, easily obtained measures in our analyses. However, this may also be considered as a weakness, as many potentially relevant variables were not available for inclusion in our analyses, including

476 medication use, treatment history, therapist competence, and therapeutic alliance, as
477 well as other key risk and protective factors. The absence of such prognostic variables
478 can be seen in the relatively low R^2 values for our models.

479 We acknowledge additional limitations in terms of sample selection and available
480 variables. The selection of referrals meeting IAPT service criteria for treatment may
481 have introduced Berkson's bias into our analyses, particularly in those regarding
482 recovery. We were unable to quantify this bias because we have no recovery data for
483 referrals not admitted to IAPT. Furthermore, the subsample of 8,114 with the requisite
484 data for evaluation of predictors of recovery may not be representative of the full sample
485 as patients drop out for non-random reasons; again, this is inherent in many clinical
486 samples where drop outs may be due to recovery or worsening. Therefore, inferences
487 from these findings need to be made with caution. Finally, whilst our sample size was
488 large, it represented only two mental health trusts, which may limit the generalisability of
489 our results.

490 In terms of limitations relating to individual variables,, we could not investigate
491 relationship between ethnicity and treatment response trajectories as we had very few
492 Black, Asian, and Minority Ethnic (BAME) individuals in our sample (though this broadly
493 reflects the proportion of BAME people accessing IAPT nationally). Whilst beyond the
494 scope of this paper, the low proportion of BAME individuals accessing IAPT services
495 merits careful consideration. Low participation may be due to a number of causes,
496 including individual factors (e.g. personal attitudes toward services), service-level
497 factors (e.g. inaccessibility or unacceptability), and wider cultural issues (e.g.
498 discrimination and stigma). Furthermore, our calculation of socioeconomic status has

limitations; as IAPT does not collect this information on an individual level, we derived this variable by using the outward code of each individual's postcode to collect the median IMD decile for the area. Finally, in terms of caseload, we were only able to identify the number of patients the therapist sees *within the IAPT setting*. Yet, it is not uncommon for IAPT therapists to see additional patients outside of IAPT.

Conclusions

Therapist confidence and self-efficacy are important factors for determining therapy effectiveness (Green, Barkham, Kellett, & Saxon, 2014; Heinonen, Lindfors, Laaksonen, & Knekt, 2012). Equally as important, patients' positive *expectations* of therapy outcome have been consistently linked to better *actual* outcomes (Mondloch, Cole, & Frank, 2001). Whilst therapists and patients may worry about progress and outcomes in the context of more severe baseline symptomatology, our findings suggest that they can take courage in the knowledge that more 'unwell' patients actually have the potential to improve more rapidly. Moreover, this finding demonstrates the gains possible for patients with more severe depression and anxiety in services offering short-term psychological therapies.

Furthermore, our results about variables associated with symptomatic recovery are useful for highlighting groups of patients that may benefit from additional or more intensive intervention. One such group consists of patients who begin therapy with more severe depression and/or anxiety symptoms and poorer functioning, as these two characteristics have a significant negative impact on symptomatic recovery. In order to

ensure that everyone has the potential to reach the IAPT symptomatic recovery threshold, we must continue to investigate how to best tailor interventions to fit individual patients' needs.

Acknowledgements

We are extremely grateful to IAPT service for providing us data. We thank Dr James Kirkbride for his expert advice on how to derive socioeconomic status from the limited information in the data available to us.

Conflicts of Interest

JS discloses consultancy for IESO digital health. The remaining authors have no conflicts of interest.

References

Ali, S., Littlewood, E., McMillan, D., Delgadillo, J., Miranda, A., Croudace, T., & Gilbody, S. (2014). Heterogeneity in Patient-Reported Outcomes following Low-Intensity Mental Health Interventions: A Multilevel Analysis. *PLoS One, 9*(9), e99658. doi:10.1371/journal.pone.0099658

Amati, F., Banks, C., Greenfield, G., & Green, J. (2018). Predictors of outcomes for patients with common mental health disorders receiving psychological therapies in community settings: a systematic review. *Journal of Public Health (Oxford, England), 40(3),* e375–e387. doi:10.1093/pubmed/fdx168

Barth, J., Munder, T., Gerger, H., Nüesch, E., Trelle, S., Znoj, H., . . . Cuijpers, P. (2013). Comparative Efficacy of Seven Psychotherapeutic Interventions for Patients with Depression: A Network Meta-Analysis. *PLoS Medicine, 10*(5), e1001454. doi:10.1371/journal.pmed.1001454

Bucher, M. A., Suzuki, T., & Samuel, D. B. (2019). A meta-analytic review of personality traits and their associations with mental health treatment outcomes. *Clinical Psychology Review, 70*(July 2018), 51-63. doi:10.1016/j.cpr.2019.04.002

Cameron, I. M., Crawford, J. R., Lawton, K., & Reid, I. C. (2008). Psychometric comparison of PHQ-9 and HADS for measuring depression severity in primary care. *British Journal of General Practice, 58*(546), 32-36. doi:10.3399/bjgp08X263794

Clark, D. M. (2018). Realizing the Mass Public Benefit of Evidence-Based Psychological Therapies: The IAPT Program. *Annual Review of Clinical Psychology, 14*, 159-183. doi:10.1146/annurev-clinpsy-050817-084833

Clark, D. M., Canvin, L., Green, J., Layard, R., Pilling, S., & Janecka, M. (2018). Transparency about the outcomes of mental health services (IAPT approach): an analysis of public data. *Lancet, 391*(10121), 679-686. doi:10.1016/S0140-6736(17)32133-5

Comninos, A., & Grenyer, B. F. S. (2007). The influence of interpersonal factors on the speed of recovery from major depression. *Psychotherapy Research, 17*(2), 239-249. doi:10.1080/10503300600849140

Cristea, I. A., Gentili, C., Cotet, C. D., Palomba, D., Barbui, C., & Cuijpers, P. (2017). Efficacy of psychotherapies for borderline personality disorder: A systematic review and meta-analysis. *JAMA Psychiatry*, *74*(4), 319–328. https://doi.org/10.1001/jamapsychiatry.2016.4287

Cuijpers, P., Cristea, I. A., Karyotaki, E., Reijnders, M., & Huibers, M. J. H. (2016). How effective are cognitive behavior therapies for major depression and anxiety disorders? A meta-analytic update of the evidence. *World Psychiatry : official journal of the World Psychiatric Association (WPA)*, *15*(3), 245–258. https://doi.org/10.1002/wps.20346

Cuijpers, P., Karyotaki, E., Weitz, E., Andersson, G., Hollon, S. D., & Van Straten, A. (2014). The effects of psychotherapies for major depression in adults on remission, recovery and improvement: A meta-analysis. *Journal of Affective Disorders*, *159*, 118–126. https://doi.org/10.1016/j.jad.2014.02.026

Cuijpers, P., Van Lier, P. A. C., Van Straten, A., & Donker, M. (2005). Examining differential effects of psychological treatment of depressive disorder: An

application of trajectory analyses. *Journal of Affective Disorders, 89*(1-3), 137-146. doi:10.1016/j.jad.2005.09.001

Cuijpers, P., Weitz, E., Twisk, J., Kuehner, C., Cristea, I., David, D., . . . Hollon, S. D. (2014). Gender as Predictor and Moderator of Outcome in Cognitive Behavior Therapy and Pharmacotherapy for Adult Depression: An "Individual Patient Data" Meta-Analysis. *Depression and Anxiety, 31*(11), 941-951. doi:10.1002/da.22328

Del Re, A. C., Flückiger, C., Horvath, A. O., Symonds, D., & Wampold, B. E. (2012). Therapist effects in the therapeutic alliance-outcome relationship: A restricted-maximum likelihood meta-analysis. *Clinical Psychology Review, 32*(7), 642-649. doi:10.1016/j.cpr.2012.07.002

Dixon, L. B., Holoshitz, Y., & Nossel, I. (2016). Treatment engagement of individuals experiencing mental illness: review and update. *World Psychiatry: official journal of the World Psychiatric Association, 15*(1), 13-20. doi:10.1002/wps.20306

Elliott, P., Biddle, D., Hawthorne, G., Forbes, D., & Creamer, M. (2005). Patterns of treatment response in chronic posttraumatic stress disorder: An application of latent growth mixture modeling. *Journal of Traumatic Stress, 18*(4), 303-311. doi:10.1002/jts.20041

Flückiger, C., Grosse Holtforth, M., Znoj, H. J., Caspar, F., & Wampold, B. E. (2013). Is the relation between early post-session reports and treatment outcome an epiphenomenon of intake distress and early response? A multi-predictor analysis in outpatient psychotherapy. *Psychotherapy Research, 23*(1), 1-13. doi:10.1080/10503307.2012.693773

Goddard, E., Wingrove, J., & Moran, P. (2015). The impact of comorbid personality difficulties on response to IAPT treatment for depression and anxiety. *Behaviour Research and Therapy, 73*, 1-7. doi:10.1016/j.brat.2015.07.006

Green, H., Barkham, M., Kellett, S., & Saxon, D. (2014). Therapist effects and IAPT Psychological Wellbeing Practitioners (PWPs): A multilevel modelling and mixed methods analysis. *Behaviour Research and Therapy, 63*, 43-54. doi:10.1016/j.brat.2014.08.009

Green, S. A., Honeybourne, E., Chalkley, S. R., Poots, A. J., Woodcock, T., Price, G., . . . Green, J. (2015). A retrospective observational analysis to identify patient and treatment-related predictors of outcomes in a community mental health programme. *BMJ Open, 5*(5), e006103. doi:10.1136/bmjopen-2014-006103

Gunn, J., Densley, K., Middleton, A., Ambresin, G., Dowrick, C., Hegarty, K., . . . Herrman, H. (2013). A trajectory-based approach to understand the factors associated with persistent depressive symptoms in primary care. *Journal of Affective Disorders, 148*(2-3), 338-346. doi:10.1016/j.jad.2012.12.021

Gyani, A., Shafran, R., Layard, R., & Clark, D. M. (2013). Behaviour Research and Therapy Enhancing recovery rates : Lessons from year one of IAPT. *Behaviour Research and Therapy, 51*(9), 597-606. doi:10.1016/j.brat.2013.06.004

Gyani, A., Shafran, R., Layard, R., & Clark, D. M. (2013). Enhancing recovery rates: lessons from year one of IAPT. *Behaviour Research and Therapy, 51*(9), 597-606. doi:10.1016/j.brat.2013.06.004

Heinonen, E., Lindfors, O., Laaksonen, M. A., & Knekt, P. (2012). Therapists' professional and personal characteristics as predictors of outcome in short- and

626 long-term psychotherapy. *Journal of Affective Disorders, 138*(3), 301-312.
627 doi:10.1016/j.jad.2012.01.023
628 Hodgekins, J., Birchwood, M., Christopher, R., Marshall, M., Coker, S., Everard, L., . . .
629 Fowler, D. (2015). Investigating trajectories of social recovery in individuals with
630 first-episode psychosis: a latent class growth analysis. *British Journal of*
631 *Psychiatry, 207*(6), 536-543. doi:10.1192/bjp.bp.114.153486
632 Horvath, A. O., Del Re, A. C., Flückiger, C., & Symonds, D. (2011). Alliance in Individual
633 Psychotherapy. *Psychotherapy, 48*(1), 9-16. doi:10.1037/a0022186
634 Knight, C., Russo, D. A., Stochl, J., Croudace, T. J., Fowler, D., Grey, N., . . . Perez, J.
635 (2020). *Prevalence of and Recovery from Common Mental Disorder including*
636 *Psychotic Experiences in the UK Primary Care Improving Access to*
637 *Psychological Therapies (IAPT) Programme*. *Journal of Affective Disorders*, *272*,
638 84–90. https://doi.org/10.1016/j.jad.2020.04.015
639 Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9 - Validity of a brief
640 depression severity measure. *Journal of General Internal Medicine, 16*(9), 606-
641 613. doi:DOI 10.1046/j.1525-1497.2001.016009606.x
642 Laaksonen, M. A., Knekt, P., & Lindfors, O. (2013). Psychological predictors of the
643 recovery from mood or anxiety disorder in short-term and long-term
644 psychotherapy during a 3-year follow-up. *Psychiatry Research, 208*(2), 162-173.
645 doi:10.1016/j.psychres.2012.09.053
646 Laaksonen, M. A., Knekt, P., Sares-Jäske, L., & Lindfors, O. (2013). Psychological
647 predictors on the outcome of short-term psychodynamic psychotherapy and
648 solution-focused therapy in the treatment of mood and anxiety disorder.
649 *European Psychiatry, 28*(2), 117-124. doi:10.1016/j.eurpsy.2011.12.002
650 Lambert, M. J. (2013). Outcome in psychotherapy: The past and important advances.
651 *Psychotherapy (Chicago, Ill.)*, *50*(1), 42–51. https://doi.org/10.1037/a0030682
652 Lindfors, O., Ojanen, S., Jääskeläinen, T., & Knekt, P. (2014). Social support as a
653 predictor of the outcome of depressive and anxiety disorder in short-term and
654 long-term psychotherapy. *Psychiatry Research, 216*(1), 44-51.
655 doi:10.1016/j.psychres.2013.12.050
656 Lutz, W., Hofmann, S. G., Rubel, J., Boswell, J. F., Shear, M. K., Gorman, J. M., . . .
657 Barlow, D. H. (2014). Patterns of Early Change and Their Relationship to
658 Outcome and Early Treatment Termination in Patients With Panic Disorder.
659 *Journal of Consulting and Clinical Psychology, 82*(2), 287-297.
660 doi:10.1037/a0035535
661 Lutz, W., Martinovich, Z., Lyons, J. S., Leon, S. C., & Stiles, W. B. (2007). Therapist
662 effects in outpatient psychotherapy: A three-level growth curve approach. *Journal*
663 *of Counseling Psychology, 54*(1), 32. doi:10.1037/0022-0167.54.1.32
664 Lutz, W., Stulz, N., & Kock, K. (2009). Patterns of early change and their relationship to
665 outcome and follow-up among patients with major depressive disorders. *Journal*
666 *of Affective Disorders, 118*(1-3), 60-68. doi:10.1016/j.jad.2009.01.019
667 Manea, L., Gilbody, S., & McMillan, D. (2012). Optimal cut-off score for diagnosing
668 depression with the Patient Health Questionnaire (PHQ-9): a meta-analysis.
669 *CMAJ, 184*(3), E191-196. doi:10.1503/cmaj.110829
670 Marttunen, M., Välikoski, M., Lindfors, O., Laaksonen, M. A., & Knekt, P. (2008).
671 Pretreatment clinical and psychosocial predictors of remission from depression

after short-term psychodynamic psychotherapy and solution-focused therapy: A 1-year follow-up study. *Psychotherapy Research, 18*(2), 191-199. doi:10.1080/10503300701429958

Mondloch, M. V., Cole, D. C., & Frank, J. W. (2001). Does how you do depend on how you think you'll do? A systemic review of the evidence for a relation between patients' recovery expectations and health outcomes. *Canadian Medical Association Journal, 165*(10), 1303-1303.

Mundt, J. C., Marks, I. M., Shear, M. K., & Greist, J. H. (2002). The Work and Social Adjustment Scale: a simple measure of impairment in functioning. *British Journal of Psychiatry, 180*, 461-464. doi:10.1192/bjp.180.5.461

Muthén, L. K., & Muthén, B. O. (1998-2019). *Mplus User's Guide. Eighth Edition*. Los Angeles, CA: Muthén & Muthén.

National Collaborating Centre for Mental Health. (2018). *The Improving Access to Psychological Therapies Manual*. Retrieved from https://www.england.nhs.uk/publication/the-improving-access-to-psychological-therapies-manual/

Nissen-Lie, H. A., Monsen, J. T., Ulleberg, P., & Rønnestad, M. H. (2013). Psychotherapists' self-reports of their interpersonal functioning and difficulties in practice as predictors of patient outcome. *Psychotherapy Research, 23*(1), 86-104. doi:10.1080/10503307.2012.735775

Norcross, J. C., & Wampold, B. E. (2011). What works for whom: Tailoring psychotherapy to the person. *Journal of Clinical Psychology, 67*(2), 127-132. doi:10.1002/jclp.20764

Porter, E., & Chambless, D. L. (2015). A systematic review of predictors and moderators of improvement in cognitive-behavioral therapy for panic disorder and agoraphobia. *Clinical Psychology Review, 42*, 179-192. doi:10.1016/j.cpr.2015.09.004

Public Health England. (2019). Public Health Profiles. Retrieved from https://fingertips.phe.org.uk/search/IAPT

R Core Team. (2019). R: A language and environment for statistical computing.R Foundation for Statistical Computing. Vienna, Austria. Retrieved from https://www.R-project.org

Robinson, L., Kellett, S., & Delgadillo, J. (2020). Dose-response patterns in low and high intensity cognitive behavioral therapy for common mental health problems. *Depression and Anxiety, 37*(3), 285-294. doi:10.1002/da.22999

Saxon, D., Firth, N., & Barkham, M. (2017). The Relationship Between Therapist Effects and Therapy Delivery Factors: Therapy Modality, Dosage, and Non-completion. *Administration and Policy in Mental Health and Mental Health Services Research, 44*(5), 705-715. doi:10.1007/s10488-016-0750-5

Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Lowe, B. (2006). A brief measure for assessing generalized anxiety disorder - The GAD-7. *Archives of Internal Medicine, 166*(10), 1092-1097. doi:DOI 10.1001/archinte.166.10.1092

Stein, N. R., Dickstein, B. D., Schuster, J., Litz, B. T., & Resick, P. A. (2012). Trajectories of Response to Treatment for Posttraumatic Stress Disorder. *Behavior Therapy, 43*(4), 790–800.. doi:10.1016/j.beth.2012.04.003

Stochl, J., Jones, P. B., Plaistow, J., Reininghaus, U., Priebe, S., Perez, J., & Croudace, T. J. (2014). Multilevel ordinal factor analysis of the positive and negative syndrome scale (PANSS). *International Journal of Methods in Psychiatric Research, 23*(1), 25-35. doi:10.1002/mpr.1429

Stochl, J., Khandaker, G. M., Lewis, G., Perez, J., Goodyer, I. M., Zammit, S., . . . Jones, P. B. (2015). Mood, anxiety and psychotic phenomena measure a common psychopathological factor. *Psychological Medicine, 45*(7), 1483-1493. doi: 10.1017/S003329171400261X.

Stulz, N., Lutz, W. G., Leach, C., Lucock, M., & Barkham, M. (2007). Shapes of early change in psychotherapy under routine outpatient conditions. *Journal of consulting and clinical psychology, 75*(6), 864-874. doi:10.1037/0022-006x.75.6.864

Sunderland, M., Wong, N. R., Hilvert-Bruce, Z., & Andrews, G. (2012). Investigating trajectories of change in psychological distress amongst patients with depression and generalised anxiety disorder treated with internet cognitive behavioural therapy. *Behaviour Research and Therapy, 50*(6), 374-380. doi:10.1016/j.brat.2012.03.005

Tiemens, B., Kloos, M., Spijker, J., Ingenhoven, T., Kampman, M., & Hendriks, G. J. (2019). Lower versus higher frequency of sessions in starting outpatient mental health care and the risk of a chronic course; a naturalistic cohort study. *BMC Psychiatry, 19*(1), 228. doi:10.1186/s12888-019-2214-4

Titov, N., Dear, B. F., McMillan, D., Anderson, T., Zou, J., & Sunderland, M. (2011). Psychometric comparison of the PHQ-9 and BDI-II for measuring response during treatment of depression. *Cognitive Behavioral Therapy, 40*(2), 126-136. doi:10.1080/16506073.2010.550059

Van, H. L., Dekker, J., Peen, J., Van Aalst, G., & Schoevers, R. A. (2008). Identifying patients at risk of complete nonresponse in the outpatient treatment of depression. *Psychotherapy and Psychosomatics, 77*(6), 358-364. doi:10.1159/000151389

Van, H. L., Schoevers, R. A., Kool, S., Hendriksen, M., Peen, J., & Dekker, J. (2008). Does early response predict outcome in psychotherapy and combined therapy for major depression? *Journal of Affective Disorders, 105*(1-3), 261-265. doi:10.1016/j.jad.2007.04.016

Wang, J. Y., Mann, F., Lloyd-Evans, B., Ma, R. M., & Johnson, S. (2018). Associations between loneliness and perceived social support and outcomes of mental health problems: a systematic review. *BMC psychiatry*, *18*(1), 156. https://doi.org/10.1186/s12888-018-1736-5

Wolitzky-Taylor, K. B., Arch, J. J., Rosenfield, D., & Craske, M. G. (2012). Moderators and non-specific predictors of treatment outcome for anxiety disorders: A comparison of cognitive behavioral therapy to acceptance and commitment therapy. *Journal of Consulting and Clinical Psychology, 80*(5), 786-799. doi:10.1037/a0029418

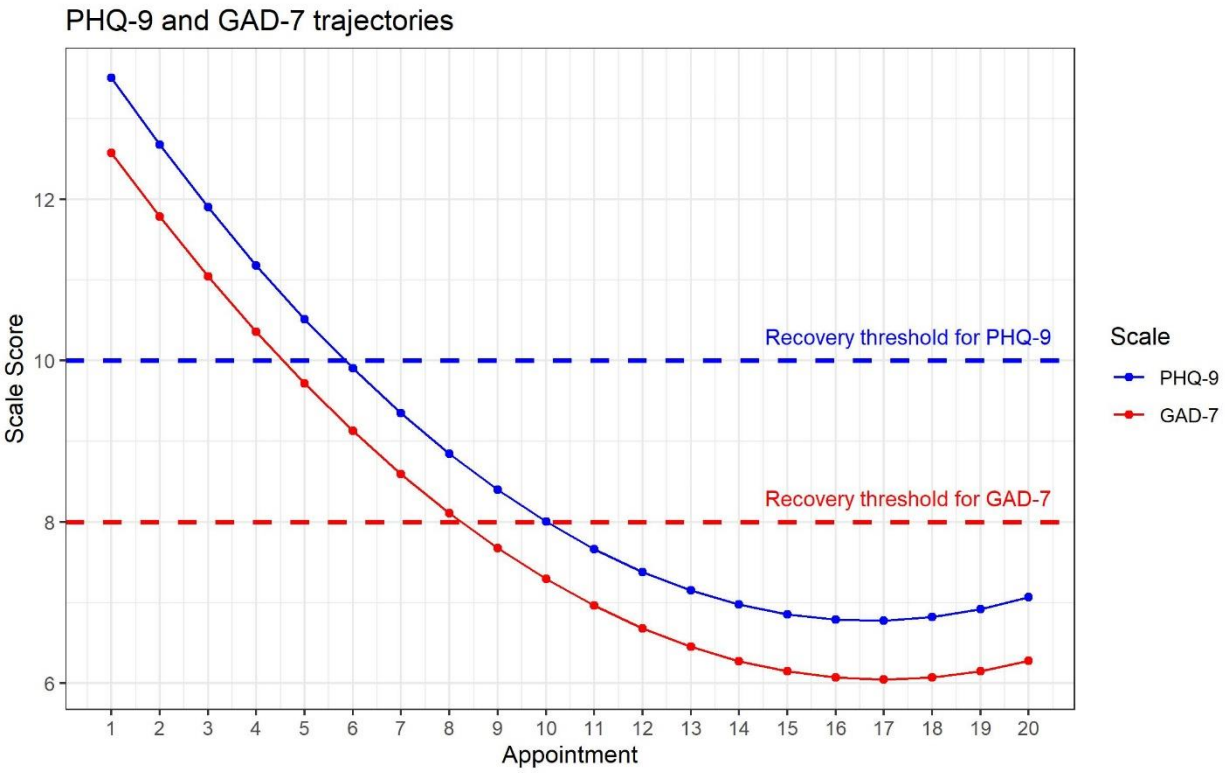


Figure 1: Estimated average growth model trajectories for PHQ-9 (blue) and GAD-7 (red). Scale score is computed as sum score of all items.

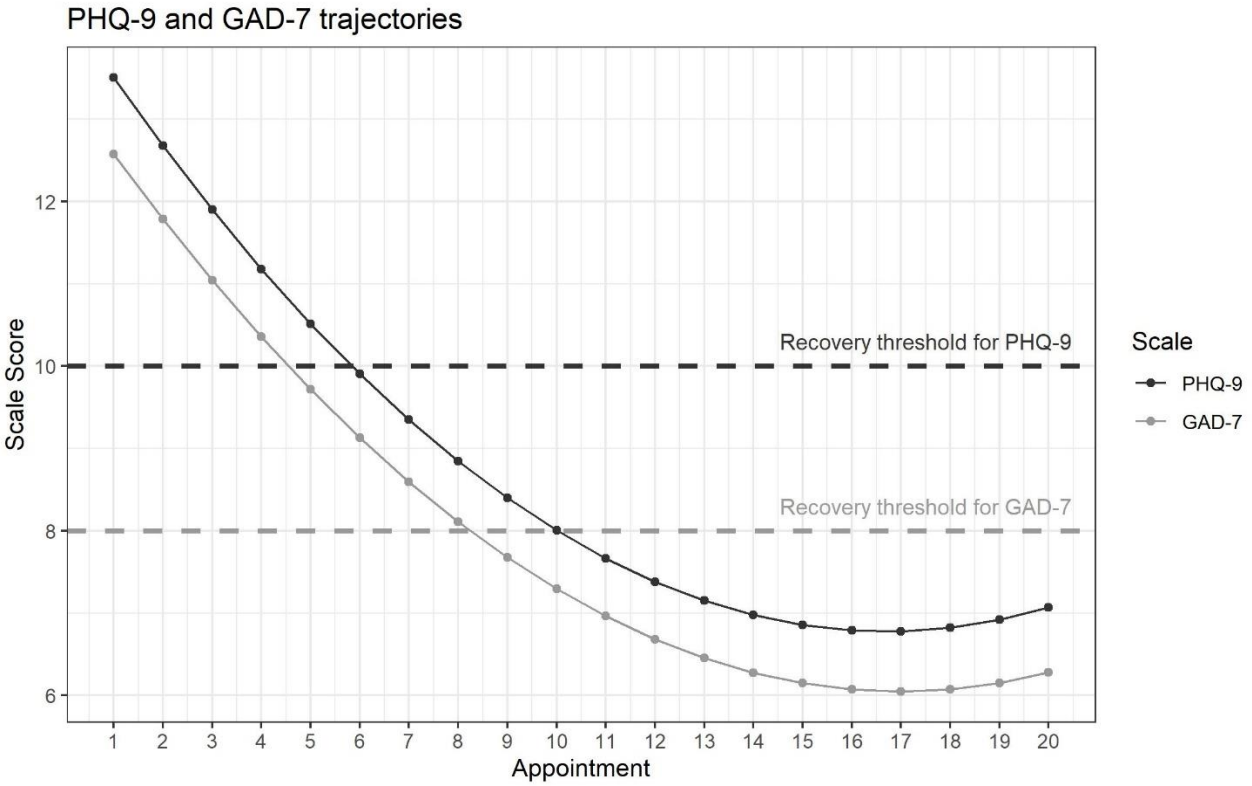


Figure 1: Estimated average growth model trajectories for PHQ-9 (blue) and GAD-7 (red). Scale score is computed as sum score of all items.

Table 1: Sample characteristics and descriptive statistics for variables hypothesised to be associated with outcomes

	Full sample (n=28,810)	**Recovery subsample (n=8,114)**
Gender count (%)	females=19,191 (66.6) males=9,462 (32.8) missing=157 (0.6)	females=5,378 (66.3) males=2,736 (33.7) missing=0 (0)
Age mean (sd)	41.3 (15.3)	39.4 (13.3)
Functioning (WSAS): mean (sd)	19.2 (9.1)	20.0 (8.6)
Socioeconomic status (median IMD): mean (sd)	6.3 (2.0)	6.4 (1.9)
Therapy intensity: count (%)	low=17,189 (59.7)* high=10,884 (37.8)* missing=737 (2.5)	low=5,037 (62.1) high=3,077 (37.9) missing=0 (0)
Therapy frequency (gap between therapies in days): median (IQR)	14.0 (12.5)	14.0 (10)
Baseline symptom severity PHQ-9: mean (sd)	13.6 (6.3)	14.7 (5.3)
Baseline symptom severity GAD-7: mean (sd)	12.6 (5.3)	14.0 (4.1)
Number of patients per therapist: mean (sd)	54.6 (67.5)	67.9 (70.0)
Ethnicity: count (%)	White=25,254 (87.6) Indian=334 (1.2) Asian=238 (0.8) Black=206 (0.7) Mixed/other=717 (2.5) missing=2,061 (7.2)	White=6,538 (80.6) Indian=131 (1.6) Asian=86 (1.1) Black=63 (0.8) Mixed/other=235 (2.9) missing=1,061 (13.0)
Diagnosis count (%)	Depression= 11,194 (38.9) Anxiety=9,219 (32.0) OCD=839 (2.9) PTSD=1,166 (4.0) Panic disorder=791 (2.7) Phobia=1,084 (3.8) Other/unspecified=636 (2.2) missing=3,881 (13.5)	Depression=3,096 (38.2) Anxiety=2,562 (31.6) OCD=270 (3.3) PTSD=304 (3.8) Panic disorder=252 (3.1) Phobia=294 (3.6) Other/unspecified=239 (2.9) missing=1,097 (13.5)

* Of those initially assigned to low or high intensity therapy 9,544 individuals (33.1%) were stepped up or down in intensity during the therapy course.

OCD=Obsessive-compulsive disorder; PTSD=Post-traumatic stress disorder

Table 2: Regression coefficients for conditional multilevel growth model of depressive symptoms

Level	Outcome	Independent variable*	Estimate (standard error)	Standardised estimate (standard error)	p-value
Individual level (within)	BSS	Gender[a]	-0.401 (0.081)	-0.034 (0.007)	<0.001
	BSS	Age	-0.021 (0.003)	-0.051 (0.007)	<0.001
	BSS	Functioning (WSAS)	0.337 (0.005)	0.547 (0.007)	<0.001
	BSS	Socioeconomic status (median IMD)	-0.177 (0.026)	-0.061 (0.009)	<0.001
	BSS	Therapy intensity	0.374 (0.103)	0.032 (0.009)	<0.001
	RoSCh	Gender[a]	-0.190 (0.073)	-0.025 (0.010)	0.009
	RoSCh	Age	-0.017 (0.003)	-0.063 (0.011)	<0.001
	RoSCh	Functioning (WSAS)	0.022 (0.005)	0.055 (0.013)	<0.001
	RoSCh	Socioeconomic status (median IMD)	0.068 (0.032)	0.037 (0.017)	0.037
	RoSCh	Therapy intensity	-0.005 (0.090)	-0.001 (0.012)	0.960
	RoSCh	Therapy frequency	0.002 (0.006)	0.010 (0.026)	0.681
	RoSCh	Baseline symptom severity (Intercepts within)	-0.252 (0.013)	-0.393 (0.014)	<0.001
Therapist level (between)	RoSCh	Number of patients	-0.005 (0.003)	-0.319 (0.153)	0.037

* All independent variables are cross-adjusted

[a] Reference group=females

RoSCh = Rate of symptomatic change (where more negative slopes indicate more rapid improvement), BSS=Baseline Symptom Severity

R-squares: RoSCh=0.138, BSS=0.321

Table 3: Regression coefficients for conditional multilevel growth model of anxiety symptoms

Level	Outcome	Independent variable*	Estimate (standard error)	Standardised estimate (standard error)	p-value
Individual level (within)	BSS	Gender[a]	-0.937 (0.072)	-0.095 (0.007)	<0.001
	BSS	Age	-0.036 (0.003)	-0.102 (0.008)	<0.001
	BSS	Functioning (WSAS)	0.207 (0.004)	0.403 (0.008)	<0.001
	BSS	Socioeconomic status (median IMD)	-0.115 (0.024)	-0.048 (0.010)	<0.001
	BSS	Therapy intensity	-0.193 (0.097)	-0.020 (0.010)	0.048
	RoSCh	Gender[a]	-0.092 (0.063)	-0.014 (0.010)	0.145
	RoSCh	Age	-0.014 (0.003)	-0.058 (0.010)	<0.001
	RoSCh	Functioning (WSAS)	0.044 (0.004)	0.126 (0.012)	<0.001
	RoSCh	Socioeconomic status (median IMD)	0.013 (0.024)	0.008 (0.015)	0.577
	RoSCh	Therapy intensity	0.005 (0.088)	0.001 (0.013)	0.955
	RoSCh	Therapy frequency	0.008 (0.005)	0.042 (0.025)	0.091
	RoSCh	Baseline symptom severity (Intercepts within)	-0.262 (0.012)	-0.388 (0.012)	<0.001
Therapist level (between)	RoSCh	Number of patients	-0.006 (0.002)	-0.322 (0.092)	0.001

* All independent variables are cross-adjusted

[a] Reference group=females

RoSCh = Rate of symptomatic change (where more negative slopes indicate more rapid improvement), BSS=Baseline Symptom Severity

R-squares: RoSCh=0.127, BSS=0.188

Table 4: Regression coefficients, odds ratios and bootstrapped odds ratios for variables hypothesised to be associated with symptomatic recovery

Level	Independent variable*	Estimate (standard error)	Odds ratio	p-value	Bootstrapped odds ratios mean (95% CI)
Individual level (within)	Baseline symptom severity PHQ-9	-0.060 (0.007)	0.942	<0.001	0.940 (0.922-0.958)
	Baseline symptom severity GAD-7	-0.065 (0.007)	0.937	<0.001	0.935 (0.916-0.953)
	Gender	0.076 (0.049)	1.079	0.138	1.090 (0.947-1.248)
	Age	0.008 (0.002)	1.008	<0.001	1.008 (1.003-1.014)
	Functioning (WSAS)	-0.024 (0.003)	0.976	<0.001	0.975 (0.966-0.985)
	Therapy frequency	-0.008 (0.003)	0.992	0.015	0.994 (0.985-1.003)
	Therapy intensity	-0.119 (0.060)	0.888	0.034	0.886 (0.748-1.048)
	Socioeconomic status (median IMD)	-0.021 (0.016)	0.979	0.180	0.985 (0.941-1.033)
Therapist level (between)	Number of patients	0.000 (0.001)	1.000	0.446	1.001 (0.999-1.002)

* All independent variables were cross-adjusted

Appendix 1: Classes of treatment response trajectories.

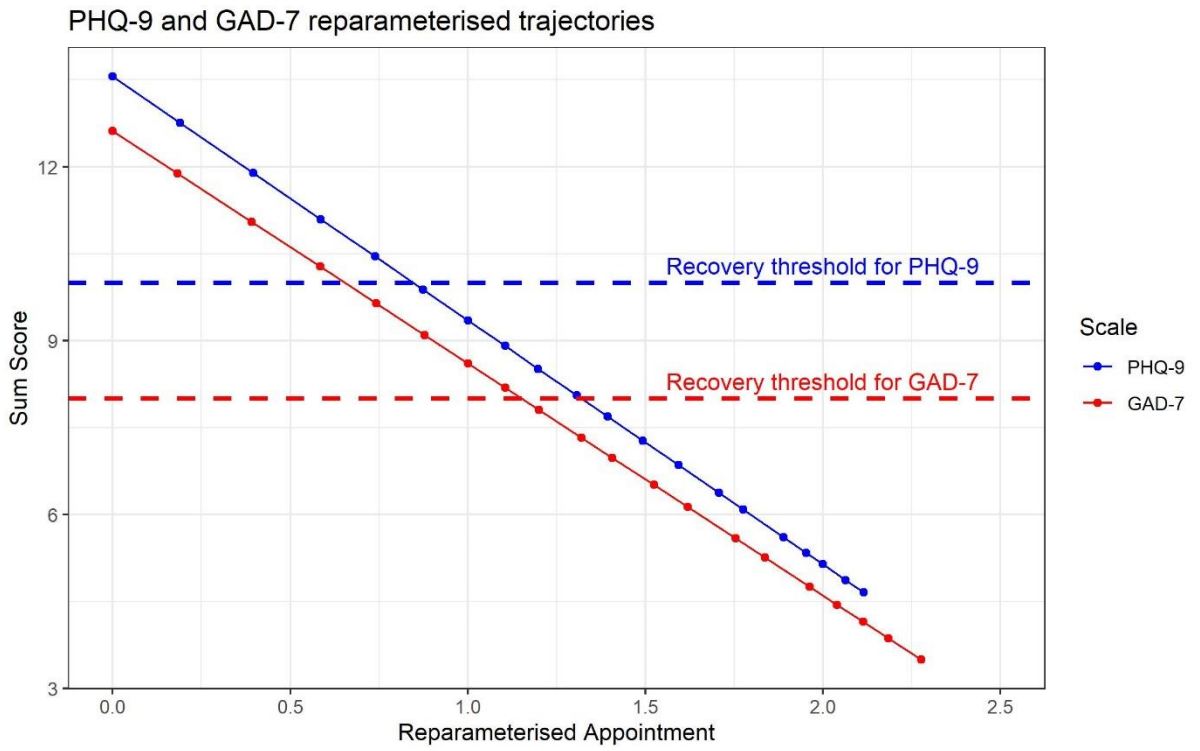
A growth mixture model with 1-9 classes was fitted to the data to establish if treatment response trajectories cluster into interpretable homogeneous classes. The fit of the estimated models is presented in Supplementary Table 1. Dropping indices such as Bayesian Information Criterion (BIC), Akaike Information Criterion (AIC) in combination with entropy values do not support few homogeneous, yet distinct classes.

Supplementary Table 1: Fit of growth mixture model with 1-9 classes.

	Title	Observations	Number of parameters	Log-Likelihood	AIC	BIC	Entropy
PHQ-9	1-class	27835	23	-555535	1111117	1111306	-
	2-classes	27835	27	-514358	1028771	1028993	0.828
	3-classes	27835	31	-498217	996497	996752	0.801
	4-classes	27835	35	-491410	982890	983178	0.773
	5-classes	27835	39	-488228	976534	976855	0.743
	6-classes	27835	43	-485876	971839	972193	0.723
	7-classes	27835	47	-484107	968307	968694	0.712
	8-classes	27835	51	-482741	965585	966005	0.688
	9-classes	27835	55	-481639	963387	963840	0.681
GAD-7	1-class	27832	23	-531038	1062122	1062311	-
	2-classes	27832	27	-493235	986523	986745	0.816
	3-classes	27832	31	-479348	958757	959012	0.78
	4-classes	27832	35	-473553	947177	947465	0.746
	5-classes	27832	39	-471127	942332	942653	0.719
	6-classes	27832	43	-468636	937358	937713	0.699
	7-classes	27832	47	-467019	934131	934518	0.676
	8-classes	27832	51	-465694	931491	931911	0.669
	9-classes	27832	55	-464716	929543	929995	0.660

Appendix 2: Reparameterised model

The traditional parameterisation of growth models is useful for estimating the shape of treatment response trajectories but limits the interpretation of the slope as the average improvement in scores between the first and second therapy session when the trajectory shape is, as in our case, nonlinear. Fortunately, the model can be reparameterised such that the slope can be interpreted as growth between two arbitrarily-chosen appointments. Times for the appointments are estimated for all except for the anchoring appointments (i.e. the anchor appointments are fixed), so that the. Non-linearity of recovery trajectories is still taken into account. In our case, we have chosen the 1st and the 7th appointments as our anchoring appointments. Therefore, the slope can be interpreted as the *difference in modelled scores between these two appointments*. The result is a linear model, although the non-linearity resulting from differential changes in symptoms between appointments is still taken into account. The fit of this model is slightly worse compared to non-reparameterised model, yet still acceptable for both the PHQ-9 (RMSEA=0.042, RMSEA 90% CI=(0.042-0.043), CFI=0.923, TLI=0.922, SRMR=0.145) and the GAD-7 (RMSEA=0.042, RMSEA 90% CI=(0.041-0.043), CFI=0.929, TLI=0.922, SRMR=0.183). The treatment response trajectories of the reparameterised model are depicted in Supplementary Figure 2). The slope of the reparameterised growth model has a value of -4.208 for the PHQ-9 and of -4.005 for the GAD-7. This suggests that, on average, patients improve by about 4.2 points on PHQ-9 score, and by 4 points on GAD-7 score, over the course of 7 appointments. The correlation between slope and the intercept was negative (-0.331 for PHQ-9 and -0.311 for GAD-7), suggesting that individuals who are initially more depressed and anxious tend to improve faster on the corresponding scale.



Supplementary Figure 2: Estimated growth model trajectories for the reparameterised growth model (PHQ-9 in blue and GAD-7 in red)

Appendix 3: Multilevel modelling considerations

Here we have now investigated the amount of variance accounted for by the therapist and site levels in our outcomes of interest (slopes and intercepts of growth curves and recovery). For this we used the intraclass correlation coefficient (ICC) and design effect (DEFF) coefficient. ICCs > 0.05 (5%) and DEFFs > 2 indicate a need for multilevel modelling at the corresponding level (Lai & Kwok, 2015; Muthen & Satorra, 1995). DEFF is inappropriate for very small or very large cluster sizes (Lai & Kwok, 2015) and thus we do not report it for sites (where the average cluster size is of thousands). Results are presented in the table below and suggest that using a two-level model is appropriate for this dataset.

Supplementary Table 2: Intraclass correlation coefficients and design effects for therapist and site levels.

Level	Outcome	Intraclass correlation coefficient (% of total outcome variance)	Average cluster size	Design effect
Therapist	Slopes (rates of symptomatic change)	0.051 (5.1%)	53.73	3.67
	Intercepts (Baseline symptom severity)	0.090 (9.0%)	53.73	5.76
	Recovery	0.056 (5.6%)	19.60	2.03
Site	Slopes (rates of symptomatic change)	0.003 (0.3%)*	13,916	-
	Intercepts (Baseline symptom severity)	0.005 (0.5%)*	13,916	-
	Recovery	-**	4,057	-

* only 2 clusters (sites) available. Interpret with caution
** estimation failed – likely as a combination of binary outcome (recovery) and 2 clusters (sites)

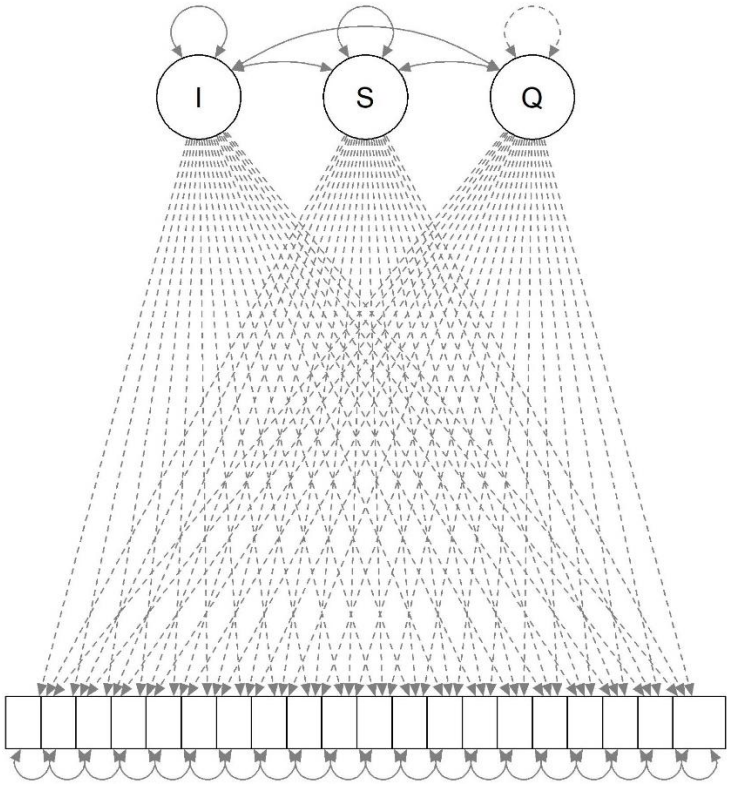
Appendix 4: Variables associated with symptomatic recovery for the more inclusive sample (n=13,349)

This additional analysis presents the results of the recalculation of our recovery model for a sample that consists of n=8,114 individuals who completed treatment (as presented in the main paper) as well as an additional n=5,235 individuals who a) had not yet completed treatment or had dropped out, b) had attended at least 3 therapy sessions, and c) had both PHQ-9 and GAD-7 scores available at their last recorded session. Their symptomatic recovery status was derived from their PHQ-9 and GAD-7 scores at the last recorded session such that they had "recovered" if both scores were below IAPT recovery thresholds (PHQ-9 < 10 and GAD-7 < 7). The regression coefficients (and odds ratios) are very similar to those presented in the main text (see Table 4). Standard errors and p-values were generally smaller in this more inclusive sample because of the larger sample size.

Supplementary Table 3: Regression coefficients, odds ratios and bootstrapped odds ratios for variables hypothesised to be associated with symptomatic recovery for the more inclusive sample (n=13,349)

Level	Independent variable*	Estimate (standard error)	Odds ratio	p-value	Bootstrapped odds ratios mean (95% CI)
Individual level (within)	Baseline symptom severity PHQ-9	-0.074 (0.005)	0.929	<0.001	0.939 (0.920-0.958)
	Baseline symptom severity GAD-7	-0.062 (0.005)	0.940	<0.001	0.935 (0.914-0.956)
	Gender	0.128 (0.039)	1.136	0.001	1.069 (0.905-1.242)
	Age	0.014 (0.001)	1.015	<0.001	1.008 (1.002-1.015)
	Functioning (WSAS)	-0.025 (0.003)	0.975	<0.001	0.975 (0.965-0.986)
	Therapy frequency	-0.031 (0.002)	0.969	<0.001	0.992 (0.982-1.001)
	Therapy intensity	-0.110 (0.050)	0.896	0.027	0.862 (0.724-1.036)
	Socioeconomic status (median IMD)	-0.015 (0.013)	0.986	0.258	0.981 (0.934-1.031)
Therapist level (between)	Number of patients	0.000 (<0.001)	1.000	0.750	1.000 (0.999-1.002)

* All independent variables were cross-adjusted



Supplementary Figure 1: Conceptual path diagram of unconditional nonlinear growth model. Squares represent observed scores (PHQ-9 or GAD-7) at first 20 appointments, ovals represent latent intercepts (I), slopes (S) and quadratic terms (Q). Dashed arrows show fixed parameters.

Supplementary Table 1: List of available variables routinely collected in IAPT and derived variables. Variables in bold were used as predictors, variables in italics were used to derive additional variables.

Available variables routinely collected in IAPT	*Derived variables*
Therapy intensity	**Age** (derived from Year and Month of birth)
PHQ-9	**IMD** (derived from Postcode)
GAD-7	**Therapy frequency** (derived from Appointment dates)
WSAS (at baseline)	**Number of patients** (derived from Therapist ID)
Gender	Recovery (derived from PHQ-9, GAD-7, Appointment number, and End of care reason)
Postcode (we had access to outward area portion of postcode)	
Therapist ID	
Appointment Date	
Appointment number	
Month of birth	
Year of birth	
End of care reason	
End of care date	
Ethnicity	
Referral date	
Site	
Service code	
Religion	
Sexuality	
Problem descriptor	
Service research	